UNITED STATES PATENT APPLICATION

CRYSTALLINE OR AMORPHOUS MEDIUM-K GATE OXIDES, Y_2O_3 AND Gd_2O_3

20040164360

INVENTORS

Kie Y. Ahn

of Chappaqua, NY

and

Leonard Forbes

of Corvallis, OR

Schwegman, Lundberg, Woessner, & Kluth, P.A.

1600 TCF Tower

121 South Eighth Street

Minneapolis, Minnesota 55402

ATTORNEY DOCKET 1303.021US1

Client Ref. No. 01-0575

Drafting Atty.: David Peterson

CRYSTALLINE OR AMORPHOUS MEDIUM-K GATE OXIDES, Y_2O_3 AND Gd_2O_3

Field of the Invention

The invention relates to semiconductor devices and device fabrication. Specifically, the invention relates to gate oxide layers of transistor devices and their method of fabrication.

Background of the Invention

In the semiconductor device industry, particularly in the fabrication of transistors, there is continuous pressure to reduce the size of devices such as transistors. The ultimate goal is to fabricate increasingly smaller and more reliable integrated circuits (ICs) for use in products such as processor chips, mobile telephones, or memory devices such as DRAMs. The smaller devices are frequently powered by batteries, where there is also pressure to reduce the size of the batteries, and to extend the time between battery charges. This forces the industry to not only design smaller transistors, but to design them to operate reliably with lower power supplies.

A common configuration of a transistor is shown in Figure 1. While the following discussion uses Figure 1 to illustrate a transistor from the prior art, one skilled in the art will recognize that the present invention could be incorporated into the transistor shown in Figure 1 to form a novel transistor according to the invention. The transistor 100 is fabricated in a substrate 110 that is typically silicon, but could be fabricated from other semiconductor materials as well. The transistor 100 has a first source/drain region 120 and a second source/drain region 130. A body region 132 is located between the first source/drain region and the second source/drain region, the body region 132 defining a channel of the transistor with a channel length 134. A gate dielectric, or gate oxide 140 is located on the body region 132 with a gate 150 located over the gate oxide. Although the gate dielectric can be formed from materials other than

oxides, the gate dielectric is typically an oxide, and is commonly referred to as a gate oxide. The gate may be fabricated from polycrystalline silicon (polysilicon) or other conducting materials such as metal may be used.

In fabricating transistors to be smaller in size and reliably operating on lower power supplies, one important design criteria is the gate oxide 140. A gate oxide 140, when operating in a transistor, has both a physical gate oxide thickness and an equivalent oxide thickness (EOT). The equivalent oxide thickness quantifies the electrical properties, such as capacitance, of a gate oxide 140 in terms of a representative physical thickness. EOT is defined as the thickness of a theoretical SiO_2 layer that describes the actual electrical operating characteristics of the gate oxide 140 in the transistor 100. For example, in traditional SiO_2 gate oxides, a physical oxide thickness may be 5.0 nm, but due to undesirable electrical effects such as gate depletion, the EOT may be 6.0 nm. A gate oxide other than SiO_2 may also be described electrically in terms of an EOT. In this case, the theoretical oxide referred to in the EOT number is an equivalent SiO_2 oxide layer. For example, SiO_2 has a dielectric constant of approximately 4. An alternate oxide with a dielectric constant of 20 and a physical thickness of 100 nm would have an EOT of approximately 20 nm (100 * (4/20)), which represents a theoretical SiO_2 gate oxide.

Lower transistor operating voltages and smaller transistors require thinner equivalent oxide thicknesses (EOTs). A problem with the increasing pressure of smaller transistors and lower operating voltages is that gate oxides fabricated from SiO_2 are at their limit with regards to physical thickness and EOT. Attempts to fabricate SiO_2 gate oxides thinner than today's physical thicknesses show that these gate oxides no longer have acceptable electrical properties. As a result, the EOT of a SiO_2 gate oxide 140 can no longer be reduced by merely reducing the physical gate oxide thickness.

Attempts to solve this problem have led to interest in gate oxides made from oxide materials other than SiO_2. Certain alternate oxides have a higher dielectric constant (k), which allows the physical thickness of a gate oxide 140 to be the same as existing SiO_2 limits or thicker, but provides an EOT that is thinner than current SiO_2 limits.

A problem that arises in forming an alternate oxide layer on the body region of a transistor is the process in which the alternate oxide is formed on the body region. Recent studies show that the surface roughness of the body region has a large effect on the electrical properties of the gate oxide, and the resulting operating characteristics of the transistor. The leakage current through a physical 1.0 nm gate oxide increases by a factor of 10 for every 0.1 increase in the root-mean-square (RMS) roughness. In forming an alternate oxide layer on the body region of a transistor, a thin layer of the alternate material to be oxidized (typically a metal) must first be deposited on the body region. Current processes for depositing a metal or other alternate layer on the body region of a transistor are unacceptable due to their effect on the surface roughness of the body region.

Figure 2a shows a surface 210 of a body region 200 of a transistor. The surface 210 in the Figure has a high degree of smoothness, with a surface variation 220. Figure 2b shows the body region 200 during a conventional sputtering deposition process stage. During sputtering, particles 230 of the material to be deposited bombard the surface 210 at a high energy. When a particle 230 hits the surface 210, some particles adhere as shown by particle 235, and other particles cause damage as shown by pit 240. High energy impacts can throw off body region particles 215 to create the pits 240. A resulting layer 250 as deposited by sputtering is shown in Figure 2c. The deposited layer/body region interface 255 is shown following a rough contour created by the sputtering damage. The surface of the deposited layer 260 also shows a rough contour due to the rough interface 255.

In a typical process of forming an alternate material gate oxide, the deposited layer 250 is oxidized to convert the layer 250 to an oxide material. Existing oxidation processes do not, however, repair the surface damage created by existing deposition methods such as sputtering. As described above, surface roughness has a large influence on the electrical properties of the gate oxide and the resulting transistor.

What is needed is an alternate material gate oxide that is more reliable at existing EOTs than current gate oxides. What is also needed is an alternate material gate oxide with an EOT thinner than conventional SiO_2. What is also needed is an alternative

material gate oxide with a smooth interface between the gate oxide and the body region. Because existing methods of deposition are not capable of providing a smooth interface with an alternate material gate oxide, what is further needed is a method of forming an alternate material gate oxide that maintains a smooth interface.

Additionally, at higher process temperatures, any of several materials used to fabricate the transistor, such as silicon, can react with other materials such as metals or oxygen to form unwanted silicides or oxides. What is needed is a lower temperature process of forming gate oxides that prevents the formation of unwanted byproduct materials.

Summary of the Invention

A method of forming a gate oxide on a surface such as a transistor body region is shown where a metal layer is deposited by thermal evaporation on the body region, the metal being chosen from a group consisting of the group IIIB elements and the rare earth series of the periodic table. The metal layer is then oxidized to convert the metal layer to a gate oxide. In one embodiment, the metal layer includes yttrium (Y). In one embodiment, the metal layer includes gadolinium (Gd). One embodiment of the invention uses an electron beam source to evaporate the metal layer onto the body region of the transistor. The oxidation process in one embodiment utilizes a krypton(Kr)/oxygen (O_2) mixed plasma process.

In addition to the novel process of forming a gate oxide layer, a transistor formed by the novel process exhibits novel features that may only be formed by the novel process. Thermal evaporation deposition of a metal layer onto a body region of a transistor preserves an original smooth surface roughness of the body region in contrast to other prior deposition methods that increase surface roughness. The resulting transistor fabricated with the process of this invention will exhibit a gate oxide/body region interface with a surface roughness variation as low as 0.6 nm.

These and other embodiments, aspects, advantages, and features of the present invention will be set forth in part in the description which follows, and in part will

become apparent to those skilled in the art by reference to the following description of the invention and referenced drawings or by practice of the invention. The aspects, advantages, and features of the invention are realized and attained by means of the instrumentalities, procedures, and combinations particularly pointed out in the appended claims.

Brief Description of the Drawings

Figure 1 shows a common configuration of a transistor.

Figure 2a shows a smooth surface of a body region of a transistor.

Figure 2b shows a deposition process according to the prior art.

Figure 2c shows a deposited film on a body region according to the prior art.

Figure 3a shows a deposition process according to the invention.

Figure 3b shows a magnified view of a deposited film on a body region from Figure 3a.

Figure 4a shows a deposited film on a body region according to the invention.

Figure 4b shows a partially oxidized film on a body region according to the invention.

Figure 4c shows a completely oxidized film on a body region according to the invention.

Figure 5 shows a perspective view of a personal computer.

Figure 6 shows a schematic view of a central processing unit.

Figure 7 shows a schematic view of a DRAM memory device.

Detailed Description of the Preferred Embodiments

In the following detailed description of the invention, reference is made to the accompanying drawings which form a part hereof, and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention. The terms wafer and substrate used in the following description include any

structure having an exposed surface with which to form the integrated circuit (IC) structure of the invention. The term substrate is understood to include semiconductor wafers. The term substrate is also used to refer to semiconductor structures during processing, and may include other layers that have been fabricated thereupon. Both wafer and substrate include doped and undoped semiconductors, epitaxial semiconductor layers supported by a base semiconductor or insulator, as well as other semiconductor structures well known to one skilled in the art. The term conductor is understood to include semiconductors, and the term insulator or dielectric is defined to include any material that is less electrically conductive than the materials referred to as conductors.

The term "horizontal" as used in this application is defined as a plane parallel to the conventional plane or surface of a wafer or substrate, regardless of the orientation of the wafer or substrate. The term "vertical" refers to a direction perpendicular to the horizonal as defined above. Prepositions, such as "on", "side" (as in "sidewall"), "higher", "lower", "over" and "under" are defined with respect to the conventional plane or surface being on the top surface of the wafer or substrate, regardless of the orientation of the wafer or substrate. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims, along with the full scope of equivalents to which such claims are entitled.

Figure 3a shows an electron beam evaporation technique to deposit a material on a surface such as a body region of a transistor. In Figure 3a, a substrate 310 is placed inside a deposition chamber 300. The substrate in this embodiment is masked by a first masking structure 312 and a second masking structure 314. In this embodiment, the unmasked region 316 includes a body region of a transistor, however one skilled in the art will recognize that other semiconductor device structures may utilize this process. Also located within the deposition chamber 300 is an electron beam source 330, and a target material 334. Although in this embodiment, an electron beam evaporation technique is used, it will be apparent to one skilled in the art that other thermal evaporation techniques can be used without departing from the scope of the invention. During the evaporation process, the electron beam source 330 generates an electron beam 332. The electron

beam hits the target material 334 and heats a portion of the target material enough to cause the surface of the target material to evaporate. The evaporated material 336 is then distributed throughout the chamber 300, and the material 336 deposits on surfaces that it contacts, such as the exposed body region 316. The depositing material builds up to form a layer 320 of material that is chemically the same as the target material 334.

In one embodiment of the invention, the deposited material layer 320 includes a pure metal layer chosen from the alkaline earth metals in group IIIB of the periodic table or the rare earth series of the periodic table. In one embodiment of the invention, the deposited material layer 320 includes yttrium (Y). In another embodiment of the invention, the deposited material layer 320 includes gadolinium (Gd). In one embodiment of the invention, the target material is a 99.9999% pure slug of yttrium. In another embodiment of the invention, the target material is a 99.9999% pure slug of gadolinium. The choices of materials were based on the properties of the oxides formed by these materials. Considerations included the thermodynamic stability of the oxide with silicon, the diffusion coefficient of the oxide at high processing temperatures such as 1000° K, the lattice match of the oxide with silicon, the dielectric constant of the oxide, and the conduction band offset of the oxide. In one embodiment, the conduction band offset of the metal oxide formed is over 2 eV. In one embodiment, the deposited material layer 320 is substantially amorphous. A lower presence of grain boundaries in the substantially amorphous material layer 320 reduces the leakage current through the final gate oxide. Although the amorphous form is preferred, the materials chosen for oxidation, such as yttrium and gadolinium are also acceptable in their crystalline form.

A thermal evaporation process such as the electron beam evaporation technique described above does not cause the surface damage that is inherent in other deposition techniques such as the sputtering technique shown in Figure 2b. This allows a very thin layer of material to be deposited on a body region of a transistor, while maintaining a smooth interface. A thermal evaporation process such as the electron beam evaporation technique described above also allows low processing temperatures that inhibit the formation of unwanted byproducts such as silicides and oxides. In one embodiment, the

thermal evaporation is performed with a substrate temperature between approximately 150 and 200 °C.

Figure 3b shows a magnified view of the body region 316 and the deposited layer 320 from Figure 3a. The interface 340 is shown with a roughness variation 346. The deposited layer surface 348 is also shown with a similar surface roughness. One possible surface variation 346 would be an atomic layer variation. In atomic smoothness, the greatest difference in surface features is between a first atomic layer as indicated by layer 342 and a second atomic layer 344. The thermal evaporation deposition technique described above preserves atomic smoothness such as is shown in Figure 3b, however other acceptable levels of surface roughness greater than atomic smoothness will also be preserved by the thermal evaporation technique.

Figures 4a-4c show a low temperature oxidation process that is used in one embodiment to convert the deposited layer 320 into a gate oxide. A deposited material layer 410 is shown in Figure 4a on a substrate surface 400. The layer 410 forms an interface 420 with the substrate surface 400, and the layer 410 has an outer surface 430. The layer 410 in this embodiment is deposited over a body region of a transistor, however the layer may be deposited on any surface within the scope of the invention.

In Figure 4b, the layer 410 is in the process of being oxidized. In one embodiment, the oxidation process includes a krypton/oxygen mixed plasma oxidation process. The mixed plasma process generates atomic oxygen or oxygen radicals in contrast to molecular oxygen or O_2 used in conventional thermal oxidation. The atomic oxygen is introduced to the layer from all exposed directions as indicated by arrows 440, creating an oxide portion 450. The atomic oxygen continues to react with the layer and creates an oxidation interface 422. As the reaction progresses, atomic oxygen diffuses through the oxide portion 450 and reacts at the oxidation interface 422 until the layer is completely converted to an oxide of the deposited material layer. Figure 4c shows the resulting oxide layer 450 which spans a physical thickness 452 from the outer surface 430 to the interface 420.

In one embodiment, the processing variables for the mixed plasma oxidation include a low ion bombardment energy of less than 7 eV, a high plasma density above $10^{12}/cm^3$ and a low electron temperature below 1.3 eV. In one embodiment, the substrate temperature is approximately 400 °C. In one embodiment, a mixed gas of 3% oxygen with the balance being krypton at a pressure of 1 Torr is used. In one embodiment, a microwave power density of 5 W/cm^2 is used.

The low temperature mixed plasma oxidation process described above allows the deposited layer to be oxidized at a low temperature, which inhibits the formation of unwanted byproducts such as silicides and oxides. The mixed plasma process in one embodiment is performed at approximately 400 °C in contrast to prior thermal oxidation processes that are performed at approximately 1000 °C. The mixed plasma oxidation process has also been shown to provide improved thickness variation on silicon (111) surfaces in addition to (100) surfaces. Although the low temperature mixed plasma process above describes the formation of alternate material oxides, one skilled in the art will recognize that the process can also be used to form SiO_2 oxide structures.

Metals chosen from group IIIB of the periodic table or the rare earth series of the periodic table form oxides that are thermodynamically stable such that the gate oxides formed will have minimal reactions with a silicon substrate or other structures during any later high temperature processing stages. Yttrium and gadolinium are two examples of metals taken from the above listed periodic table groups that form thermodynamically stable gate oxides. In particular, yttrium forms an oxide comprised of Y_2O_3 and gadolinium forms a similar oxide of Gd_2O_3. In addition to the stable thermodynamic properties inherent in the oxides chosen, the novel process used to form the oxide layers is performed at lower temperatures than the prior art, which further inhibits reactions with the silicon substrate or other structures.

A transistor made using the novel gate oxide process described above will possess several novel features. By creating an oxide material with a higher dielectric constant (k) and controlling surface roughness during formation, a gate oxide can be formed with an EOT thinner than 2 nm. A thicker gate oxide that is more uniform, and easier to process

can also be formed with the alternate material oxide of the present invention, the alternate material gate oxide possessing an EOT equivalent to the current limits of SiO_2 gate oxides. The smooth surface of the body region is preserved during processing, and a resulting transistor will have a smooth interface between the body region and the gate oxide with a surface roughness on the order of 0.6 nm.

Transistors created by the methods described above may be implemented into memory devices and information handling devices as shown in Figures 5-7 and described below. While specific types of memory devices and computing devices are shown below, it will be recognized by one skilled in the art that several types of memory devices and information handling devices could utilize the invention.

A personal computer, as shown in Figures 5 and 6, include a monitor 500, keyboard input 502 and a central processing unit 504. The processor unit typically includes microprocessor 606, memory bus circuit 608 having a plurality of memory slots 612(a-n), and other peripheral circuitry 610. Peripheral circuitry 610 permits various peripheral devices 624 to interface processor-memory bus 620 over input/output (I/O) bus 622. The personal computer shown in Figures 5 and 6 also includes at least one transistor having a gate oxide according to the teachings of the present invention.

Microprocessor 606 produces control and address signals to control the exchange of data between memory bus circuit 608 and microprocessor 606 and between memory bus circuit 608 and peripheral circuitry 610. This exchange of data is accomplished over high speed memory bus 620 and over high speed I/O bus 622.

Coupled to memory bus 620 are a plurality of memory slots 612(a-n) which receive memory devices well known to those skilled in the art. For example, single in-line memory modules (SIMMs) and dual in-line memory modules (DIMMs) may be used in the implementation of the present invention.

These memory devices can be produced in a variety of designs which provide different methods of reading from and writing to the dynamic memory cells of memory slots 612. One such method is the page mode operation. Page mode operations in a DRAM are defined by the method of accessing a row of a memory cell arrays and

randomly accessing different columns of the array. Data stored at the row and column intersection can be read and output while that column is accessed. Page mode DRAMs require access steps which limit the communication speed of memory circuit 608. A typical communication speed for a DRAM device using page mode is approximately 33 MHZ.

An alternate type of device is the extended data output (EDO) memory which allows data stored at a memory array address to be available as output after the addressed column has been closed. This memory can increase some communication speeds by allowing shorter access signals without reducing the time in which memory output data is available on memory bus 620. Other alternative types of devices include SDRAM, DDR SDRAM, SLDRAM and Direct RDRAM as well as others such as SRAM or Flash memories.

Figure 7 is a block diagram of an illustrative DRAM device 700 compatible with memory slots 612(a-n). The description of DRAM 700 has been simplified for purposes of illustrating a DRAM memory device and is not intended to be a complete description of all the features of a DRAM. Those skilled in the art will recognize that a wide variety of memory devices may be used in the implementation of the present invention. The example of a DRAM memory device shown in Figure 7 includes at least one transistor having a gate oxide according to the teachings of the present invention.

Control, address and data information provided over memory bus 620 is further represented by individual inputs to DRAM 700, as shown in Figure 7. These individual representations are illustrated by data lines 702, address lines 704 and various discrete lines directed to control logic 706.

As is well known in the art, DRAM 700 includes memory array 710 which in turn comprises rows and columns of addressable memory cells. Each memory cell in a row is coupled to a common wordline. Additionally, each memory cell in a column is coupled to a common bitline. Each cell in memory array 710 includes a storage capacitor and an access transistor as is conventional in the art.

DRAM 700 interfaces with, for example, microprocessor 606 through address lines 704 and data lines 702. Alternatively, DRAM 700 may interface with a DRAM controller, a micro-controller, a chip set or other electronic system. Microprocessor 606 also provides a number of control signals to DRAM 700, including but not limited to, row and column address strobe signals RAS and CAS, write enable signal WE, an output enable signal OE and other conventional control signals.

Row address buffer 712 and row decoder 714 receive and decode row addresses from row address signals provided on address lines 704 by microprocessor 606. Each unique row address corresponds to a row of cells in memory array 710. Row decoder 714 includes a wordline driver, an address decoder tree, and circuitry which translates a given row address received from row address buffers 712 and selectively activates the appropriate wordline of memory array 710 via the wordline drivers.

Column address buffer 716 and column decoder 718 receive and decode column address signals provided on address lines 704. Column decoder 718 also determines when a column is defective and the address of a replacement column. Column decoder 718 is coupled to sense amplifiers 720. Sense amplifiers 720 are coupled to complementary pairs of bitlines of memory array 710.

Sense amplifiers 720 are coupled to data-in buffer 722 and data-out buffer 724. Data-in buffers 722 and data-out buffers 724 are coupled to data lines 702. During a write operation, data lines 702 provide data to data-in buffer 722. Sense amplifier 720 receives data from data-in buffer 722 and stores the data in memory array 710 as a charge on a capacitor of a cell at an address specified on address lines 704.

During a read operation, DRAM 700 transfers data to microprocessor 606 from memory array 710. Complementary bitlines for the accessed cell are equilibrated during a precharge operation to a reference voltage provided by an equilibration circuit and a reference voltage supply. The charge stored in the accessed cell is then shared with the associated bitlines. A sense amplifier of sense amplifiers 720 detects and amplifies a difference in voltage between the complementary bitlines. The sense amplifier passes the amplified voltage to data-out buffer 724.

Control logic 706 is used to control the many available functions of DRAM 700. In addition, various control circuits and signals not detailed herein initiate and synchronize DRAM 700 operation as known to those skilled in the art. As stated above, the description of DRAM 700 has been simplified for purposes of illustrating the present invention and is not intended to be a complete description of all the features of a DRAM. Those skilled in the art will recognize that a wide variety of memory devices, including but not limited to, SDRAMs, SLDRAMs, RDRAMs and other DRAMs and SRAMs, VRAMs and EEPROMs, may be used in the implementation of the present invention. The DRAM implementation described herein is illustrative only and not intended to be exclusive or limiting.

Conclusion

Thus has been shown a gate oxide and method of fabricating a gate oxide that produce a more reliable and thinner equivalent oxide thickness. Gate oxides formed from elements in group IIIB of the periodic table or the rare earth series of the periodic table are thermodynamically stable such that the gate oxides formed will have minimal reactions with a silicon substrate or other structures during any later high temperature processing stages. Yttrium and gadolinium oxides in particular have been shown to provide excellent electrical and thermodynamic properties. In addition to the stable thermodynamic properties inherent in the oxides shown, the process shown is performed at lower temperatures than the prior art, which further inhibits reactions with the silicon substrate or other structures.

Transistors and higher level ICs or devices have been shown utilizing the novel gate oxide and process of formation. The higher dielectric constant (k) oxide materials shown in one embodiment are formed with an EOT thinner than 2 nm, e.g. thinner than possible with conventional SiO_2 gate oxides. A thicker gate oxide that is more uniform, and easier to process has also been shown with at EOT equivalent to the current limits of SiO_2 gate oxides. In one embodiment of the present invention, the novel gate oxide provides a conduction band offset of 2 eV or greater.

A novel process of forming a gate oxide has been shown where the surface smoothness of the body region is preserved during processing, and the resulting transistor has a smooth interface between the body region and the gate oxide with a surface roughness on the order of 0.6 nm. This solves the prior art problem of poor electrical properties such as high leakage current, created by unacceptable surface roughness.

Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement which is calculated to achieve the same purpose may be substituted for the specific embodiment shown. This application is intended to cover any adaptations or variations of the present invention. It is to be understood that the above description is intended to be illustrative, and not restrictive. Combinations of the above embodiments, and other embodiments will be apparent to those of skill in the art upon reviewing the above description. The scope of the invention includes any other applications in which the above structures and fabrication methods are used. The scope of the invention should be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

Patent Application - Docket No. 1303.021US1